

Data Quality Conference Call 4/5/01

Registration Re-export:

Re-exports need to be held until things are set up at this end. Phoenix/Billings re-exported last year and user pop counts went down for both as a result. The process of Registration data being exported periodically, with monthly mods as needed, matched here to visits that are frequently exported can result in duplicates. (Pt Reg files last Feb had fewer duplicates in it.)

Concerns: Phoenix and Billings numbers were lower than other areas (in the LNF process) (FY99 vs FY98 different by approx. 10%). Draft FY99 counts went up for other areas, but were not consistent among/within areas, not a lot of integrity in FY98 and 99.

Data center needs to prepare carefully for receipt of re-export. Our commitment is to define the current logic clearly, make it public, and have Stat Officers tell us what it should be. Then, we get the changes made and put the new logic into effect for the re-export.

Proposed actions:

1. We make sure that prior to any changes to the database, we have backup tapes for recovery if necessary.
2. Before re-export, we determine specific procedures (e.g., how to handle new files, old files).
3. Consider bringing in an outside expert (database person) to review re-export plan.
4. NPIRS provide documentation on how updates are handled now and what specifically is not working now, clarifying process limitations.
5. NPIRS develop plan to look at changes to database and verify that new exports/data are correct, to make sure no mistakes have been made. We also need reports from the old data to compare to the new data – to compare counts, etc.
6. Stat Officers and NPIRS work together to ensure new exports include what they should. Stat Officers, in past few years, have reviewed the data before it was considered complete and made public. (Determining trends from data is a future possibility once data is reliable.)

Traditionally NPIRS removes old data when new data is re-exported. So the center has no storage of old data for trend reporting, etc. or even comparing with the new data for an area to look for deltas. Old data needs to be kept in new database/warehouse design.

For the upcoming re-export of registration data, we will move old registration data to another file and back it up. Then we can re-point visits to the old data if the new data is found to be invalid. The Team and Rus Pittman will review the plan prepared by NPIRS to re-export and will use outside contract consultants if necessary.

Some concerns to be addressed for the re-export:

- Community of residence linking in RPMS presents the potential of losing thousands of communities of residence and user pops will fall. We are now taking this data from the visit if it is not in the registration data, so using pure raw data will cause counts to fall. This is an issue we must address if we are to use raw data.
- An attempt to use indexing didn't work (we export the index, not the name).

- If the data does not have the right number of field lengths, it may be thrown away at the Area. The Area keeps a log that can be checked so that fixes can be made before the data is re-exported.
- RPMS patients are always in the system, but if merged or flagged as ineligible, person won't re-export. Otherwise, everyone is re-exported. Can't guarantee that patient mods are all exported. SU unique id for patients still exists and we can research how the data center handles this to track hanging visits and patient changes. A useful report is one that shows unlinked visits.

Patricia will work with Karen and Stephanie to generate a list of data we're going to need.

Long-term: In the future, we need to export the registration data with each visit -- the algorithms would have to be changed to accommodate this new export process. Have to decide who is to do the review on the data; it should not be only NPIRS. We want customer and programming sides to decide what exports and reports should reflect/contain. We also need to include a header on the exports to include package, versions, patch, and how many records is included in the export (checksum function). Software has been completed that will audit site/area systems to say what version/patches they have but not cleared for general use—we need to set up a system here to handle this information. ORYX can now feedback to sites on what was received versus what was sent.

Logs:

Since Rus had extensive experience at ANMC in logging all accesses to data, we can work closely with him to set up procedures to log access and transactions in NPIRS data. Rus will work directly with NPIRS to establish other necessary logs on database.

Replicated DB:

NPIRS has advised us that all except for 19 records are replicated.

IV&V steps: 1) compare record counts, 2) run stats on tables, columns, frequency of data between random tables. Input files reflect new data and need to be in sync.

Next: set up access to small select group to look at data. Roger is to think about what policies, controls, access required from Area level (depends on what we can do with it and who will want to use it). Patricia will determine stat officer need access. Debra to get letter from Edna on guidelines on who gets access data outside their own area – requires permission from national headquarters, including names of all members of the Team. Cathy Federico will give us input with what forms/procedures are needed to document security issues. At some point, we'll consider who else needs to look at the data after this first cut.

Still waiting on SAS upgrade and IV&V before replicated database is considered ready. We can use the SQL toolkit to pull queries until SAS upgrade ready, as can those people who don't use SAS. While SAS/DB2 Connect not available, IBM has a DB2 push to SAS that can work in the interim. DB2 Connect client needed. Karen will check on this.

Issues log:

Patricia will review this and take it to the stat officers workgroup for input and development of a plan, then forward that to Linda Querec.

Patricia topic: Stat Officer workgroup will support work on standard and error reports – they need to prioritize them and decide what the reports should be doing (the logic) and how to work with NPIRS to implement. Final user pop ought to include final logic/process up to the point of adjustments and include all the columns of data Stat Officers need to verify that data – design of the report ought to be something they people can use to define and work with the data – could have pt care team begin design of this now.

General Topics:

Documentation: An extensive list has been requested. User pop logic not yet adequate to our needs...will set up a meeting to get more clarify, definitions on all language used, etc.

Inventory: Some items still missing among files we need for next user pop that needs to go back to FY97. We have inventory of export from around 5/99 to present and EPAT Listing from mainframe. Need to know location of export files prior to FY99 back to FY97. Have a gap between FY97 and FY99; need backup files on tapes we haven't found yet. (Data loss from March 2000, not all files were reprocessed.) Have identified what files appear to be missing. Some CD copies from tapes are missing, being recopied. Will document difference from missing data and missing format/media of existing data (that we have in some media).

User Pops: Patricia to introduce Workload reporting criteria to the stat officer workgroup meeting this morning.

Coordination: Nothing to report here.

Communication: Stan and Rus, at end of April, to HQE to talk with GPRA team re: their use of NPRIS data. Indian Health Leadership Council meeting is coming up in June. List of accomplishments on website and in project status report – Joan to update powerpoint on p: drive and populate it as we accomplish things.